DOCUMENT RESUME

ED 480 037 CG 032 610

AUTHOR	Vansickle, Timothy
TITLE	Types and Uses of Tests.
PUB DATE	2003-08-00
NOTE	13p.; In: Measuring Up: Assessment Issues for Teachers, Counselors, and Administrators; see CG 032 608.
PUB TYPE	Information Analyses (070)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Classification; Context Effect; Educational Assessment; *Evaluation Methods; *Test Use; *Testing

ABSTRACT

Describing the types and uses of tests may seem to be an easy task, but it is not as straightforward as it may first appear. Tests vary on many different characteristics, are used in many different ways, cross the typical assessment categories, and in some cases are so unique as to from a category unto themselves. This chapter explores many possible classification schemes and describes how tests may be used in several common settings. (Author)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 480 037

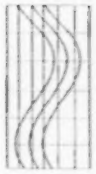
Types and Uses of Tests

By
Timothy Vansickle

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CG032610

BEST COPY AVAILABLE



Chapter 2
Types and Uses of Tests

Timothy Vansickle

Describing the types and uses of tests may seem to be an easy task, but it is not as straightforward as it may first appear. Tests vary on many different characteristics, are used in many different ways, cross the typical assessment categories, and in some cases are so unique as to form a category unto themselves. This chapter explores many possible classification schemes and describes how tests may be used in several common settings.

Types of Tests

If you open almost any textbook on psychological assessments, tests, and measurements, or any compendium of test reviews, you will find the author's classification of tests or types of tests. This classification is usually implicit in the table of contents for the book. Anastasi (1982) provides chapters or sections for individual, group, aptitude, achievement, personality, intelligence, and ability testing. Global categories include educational, occupational, and clinical, with more specific categories of self-reports, inventories, projective techniques, and so on. Janda (1998) groups tests into individual tests of intelligence, group ability tests, interests, values, structured measures of personality, projective tests and clinical assessment, neuropsychological assessment of special populations, and alternate approaches to assessment. Hopkins (1998) takes a somewhat simpler approach, with divisions into scholastic aptitude, achievement, personality, and social measures, and standardized versus instructor-made tests.

Murphy, Conoley, and Impara (1994) in the fourth edition of *Tests in Print* chose a much more linear approach to test classification, as illustrated in the following list:

- achievement
- behavior assessment
- developmental

- education
- English
- fine arts
- foreign language
- intelligence and scholastic aptitude
- math
- miscellaneous
- multi-aptitude
- neuropsychological
- personality
- reading
- science
- sensory-motor
- social studies
- speech and hearing
- vocations

As can be seen from this brief sampling, test classification is not straightforward. This confusion may result from the fact that the word *test* can be used in various ways. The new *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) defines *tests* as "all evaluative devices such as inventories [and] scales." Typical textbooks, manuscripts, and discussions use *test, assessment,* and *measure,* as well as other words, and use these interchangeably. It is, therefore, a good idea to define some of these words with a goal of enabling a classification scheme.

Allen and Yen (1979) define a test as a device for obtaining a sample of an individual's behavior. Anastasi (1982) provides a little more detail in that a test is essentially an objective and standardized measure of a sample of behavior. Hopkins (1998) suggests that a test is a technique for obtaining information. The AERA, APA, and NCME standards define a test as follows: "A test is an evaluation device or procedure in which a sample of an examinee's behavior in a specified domain is obtained and subsequently evaluated and scored using a standardized process" (p. 3).

"*Measurement* is the assigning of numbers to individuals in a systematic way as a means of representing properties of the individuals" (Allen & Yen, 1979, p. 2). Hopkins (1998) suggests that measurement is a process by which things are differentiated and described. Hence, measurement is a furthering of the testing process.

Assessment is typically the larger umbrella under which judgments,

actions, or decisions are made based on the tests and measurements used in a given situation. Assessment, therefore, includes testing and measurement, and in many contexts is used in place of either or both terms. For our discussion, we will use *test* to indicate any assessment device that might yield a score, category, or classification, or where the results could be used to make some decision about people, programs, status, or acceptance/admission.

Classifying Tests by Setting

How then do we classify tests into types or categories? Tests differ on many characteristics, such as mode of administration, stimulus materials, response mode, content, construct, level of standardization, and historical context. Test use and classification may vary with the setting in which the test is used. In clinical settings some personality tests may be classified as diagnostic while others are referred to as screening inventories. In personnel settings, tests can have a different classification system that involves selection, progression, and promotion classifications. In this setting, personality tests, aptitude tests, and achievement tests may lose their individual classifications in favor of a more global categorization such as selection battery.

Classifying Tests by Scope

One way of classifying tests may be to look at the nature of the test instrument. That is, does it have specific objectives or a narrow content domain as the target of interest? Instructor-made tests are examples of a narrowly focused type of test having specific objectives. On the other end of the continuum would be tests that measure a broad set of objectives or a large construct; for example, individually administered IQ tests. Certainly, one could argue about where on the continuum a certain type of test may fall; Figure 1 depicts one possible placement of the more general types of tests in use today.

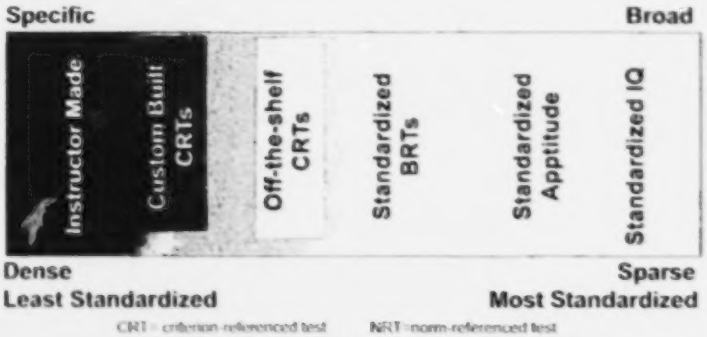


Figure 1 A test classification based on scope, number and rigor

BEST COPY AVAILABLE

In Figure 1, the number of tests also decreases as we move from left to right. Undoubtedly, there are more instructor-made tests than standardized IQ tests. Although one may argue with the placement of certain categories in Figure 1, it does provide a general sense of how tests might be classified. Additionally, Figure 1 reflects the different degrees of rigor with which tests are developed. In this regard, many instructors will argue that they standardize their tests as well as any commercial publisher, and many publishers would argue that a particular test they sell is the more rigorously developed. Some of those claims will be market driven while others are fairly subjective. Most of the broad-based intelligence tests are based on decades of research on the constructs, methods, item types, and administration procedures used. Newer, group-administered aptitude, achievement, and personality tests cannot match that history. They may however employ newer and more refined research and psychometric methods that may offset the lack of history. In presenting Figure 1, my intention is not to imply a value judgment regarding the various degrees of standardization but merely to illustrate one way of classifying tests.

It is very difficult to determine where to place cognitive tests as a group on Figure 1. For example, where does achievement end and aptitude begin? Figure 2 depicts the different overlapping possibilities in the various types of cognitive tests. Such interrelationships surely also occur in tests of personality or career interests, and in those designed for special populations. Exactly how much overlap exists is a matter of viewpoint or focus rather than a value that can be quantified empirically.

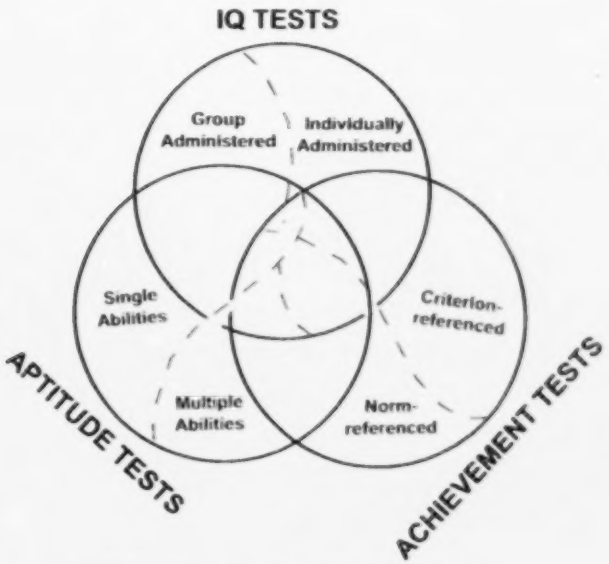


Figure 2. Interrelationships among Cognitive Tests

Classifying Tests Using a Traditional Matrix

A general classification scheme might use traditional perspectives, methodological approaches, and issues presented earlier to produce a means of classifying tests in a way useful for practitioners. Table 1 provides an example of such a matrix, including for some of the cells examples of relevant tests. Thousands of tests, inventories, and assessments are available from commercial publishers, researchers, and other practitioners. Most of these assessments are labeled as to the type of test (e.g., personality), type of administration (e.g., individual), and other characteristics and features. Although the publisher or test developer recommends certain parameters, common practice or usage may extend or restrict how an assessment is utilized, with the result that tests may overlap across cells. In addition, the practitioner could easily extend the table to include test types found most often in specific settings.

Table 1. Example Classification by Major Category, Specific Type, and Type of Administration

		Type of Administration	
Major Category/Specific Type		Group	Individual
Cognitive	Achievement	Iowa Tests of Basic Skills (1)	
		TerraNova (2)	
		Stanford9 (3)	
		The ACT Assessment (ACT) (5)	
		WorkKeys (5)	WorkKeys (5)
		Scholastic Assessment Test (SAT)(4)	
	Aptitude		Differential Aptitude Test (3)
		Cognitive Abilities Test (1)	OLSAT (3)
			Woodcock-Johnson III Tests of Cognitive Abilities (1)
	Intelligence		Stanford-Binet Intelligence Test (1)
			Wechsler Intelligence Test (3)
			Kaufman Assessment Battery for Children (K-ABC) (6)
Personality	Normal	Myers-Brigs Type Indicator (7)	16PF Fifth Edition Questionnaire (8)
		16PF Fifth Edition Questionnaire (8)	MMPI-2 (9)
			Myers-Brigs Type Indicator (7)
	Clinical	MMPI (9)	MMPI (9)
Career	Interests	Self-Directed Search (10)	Self-Directed Search (10)
		Career Decision-Making System (6)	Career Decision-Making System (6)
		Campbell Interest and Skill Survey(9)	Campbell Interest and Skill Survey(9)
	Values	Values Scale (7)	Values Scale (7)
		Career Beliefs Inventory (7)	Career Beliefs Inventory (7)
		Values Preference Indicator (11)	Values Preference Indicator (11)

(1) Riverside Publishing
(2) CTB McGraw-Hill
(3) Harcourt
(4) Educational Testing Service
(5) ACT, Inc
(6) American Guidance Service
(7) Consulting Psychologists Press
(8) Institute for Personality and Ability Testing
(9) NCS
(10) Psychological Assessment Resources
(11) Consulting Resources Group International

BEST COPY AVAILABLE

Classifying Tests by Measurement Model

A more traditional way of classifying tests is to place each test into one of several bins, including but not limited to norm-referenced versus criterion-referenced. Norm-referenced tests are those that report scores or profiles based on reference to a standard group (i.e., the norm group). People typically think of group achievement tests (e.g., Iowa Tests of Basic Skills) as belonging to this category. In addition, many personality, diagnostic, and intelligence tests also use a reference group in order to place a person into a category or to provide a score. For example, the determination of whether a client is depressed may be made in relation to a standardization group that was not depressed. In these types of tests, a normative sample of individuals is used to determine the distributional characteristics of the responses for that group (e.g., mean and standard deviation). The test is scaled so that various scores can be reported to test takers based on the typical response patterns of the standardization group. The score or scores a test taker receives are a reflection of how the person performed compared to the normative sample.

Criterion-referenced tests use a different technique to provide scores or classifications. In this case, an individual's responses are compared to some predetermined standard (i.e., criterion). The standard may be a cut-off score expressed as a raw score, a percentage, a standard score, or some other value. If the test taker reaches or exceeds the specified standard or criterion, he or she is classified as having learned the material, achieved a specific level of mastery, or falling into some group or category (e.g., addictive behavior problem).

Uses of Tests

So what have learned so far? Classification of tests can and does vary based on the classification scheme and its particular focus. Is one classification model better than another? Not necessarily. The answer depends on the purpose of the testing and the decisions one wishes to make.

Regardless of the category or classification of a test, test usage is something all practitioners must address in their work. Questions of validity, reliability, fairness, and purpose all play a part in determining the use of any instrument. Some tests may be used in multiple situations or contexts, while others may be restricted to a single situation. One key principle to remember is that a test is but a sample of an individual's behavior, learning, cognition, or other characteristic being measured.

As such, a test score should not be the sole determiner in high-stakes decisions.

What then are practitioners to do when deciding which test to use in a specific situation? First, they need to acquire training in test measurements and the specific test instrument, if required. Then, they must ask themselves a series of questions about the testing situation:

- What is the purpose of the testing?
- What decisions will be made about the person or group based on the test results?
- What tests are available for this purpose?
- Is a home-grown or a custom-built test the better option given the purpose and decisions to be made?
- What special training is required to administer and interpret the results of the test?
- What security procedures are required by either the publisher or the testing situation?
- Will the test or tests selected provide the information needed?
- Are there additional stakeholders who need different information than the test will provide?

In some cases the test user will also have to justify the cost of the testing program, in which case additional questions need to be asked:

- What is the initial purchase cost?
- What is the per-examinee cost?
- What discounts are available from the publisher (e.g., for purchasing in quantity)?
- What are the costs associated with the examinee's time (e.g., lost production time, lost instruction time)?
- What alternatives are available that might cost less?

For each context in which testing occurs, there may be additional questions that the practitioner must answer prior to selecting, administering, scoring, and interpreting a test. In the following sections, let's examine some of these particular contexts.

Testing in Schools

By far the most common situation where tests are used is in the academic setting. Whether in the K–12 or postsecondary arena, testing is a ubiquitous event in the lives of teachers, students, and administrators. Teacher-made tests to measure students' learning is by far the most prevalent form of testing. Designed well, instructor-made tests can

provide enormous amounts of information for both the teacher and the student.

In addition to teacher-made tests, many large schools and districts develop or purchase tests that they use to make decisions about the effectiveness of programs, teachers, schools, and curriculum. With the advent of the standards-based education movement, many states now incorporate statewide testing to evaluate the effectiveness of instruction and the achievement of state-established curriculum goals or targets. This typically had been done via norm-referenced tests, but standards-based initiatives have replaced or augmented the norm-referenced tests with custom-built, criterion-referenced tests designed specifically to measure the state curriculum and the success of students, teachers, programs, schools, and districts in meeting established academic targets.

Within the academic testing world, new tests are being developed to assess special populations. This is especially true with regard to statewide curriculum standards. The term *alternate assessment* is typically used to describe a test or assessment that is administered when a student's Individualized Education Program (IEP) indicates that he or she cannot be tested using the statewide test in a standard or accommodated format.

Admissions Testing

Another major area is admissions testing. The two most notable and best known of such tests are the ACT Assessment and the Scholastic Aptitude Test (SAT). The region of the country in which a student resides sometimes determines which of these two college entrance exams he or she will take. There are, of course, other admissions tests, such as the Graduate Record Exam (GRE). Most professional degree programs, such as medicine, have specialized admissions tests (e.g., the MCAT).

The goal of admissions testing is to determine who would best be served by further education in a particular field and at a particular university or college. In this respect, each school determines its own test score requirements. In the case of the ACT Assessment and SAT, the goal is to predict a particular student will be successful in the postsecondary institution to which he or she is applying. Today, however, some institutions are downplaying the importance of, or even eliminating the requirement for, a standardized college admissions test.

Tests Used in Clinical and Counseling Settings

The number and range of instruments available for use in counseling is, to say the least, staggering. Instruments exist to measure

normal personality, vocational inter[illegible], academic ability, depressive tendency, susceptibility to addictive behaviors, self-efficacy, and the need for control or dominance, to name a few. Add to these tests of intelligence or abnormal personality, plus screening and diagnostic instruments, and the practitioner in this area can quickly be inundated to the point of information overload.

Uses range from a high school counselor administering the Armed Services Vocational Aptitude Battery (ASVAB) to a clinician administering a screening instrument for depression. In these settings, the purpose of testing is to gain information about the client's characteristics or behavior. In this regard, the information may be shared with the individual for a variety of reasons, including but not limited to helping individuals make decisions about career or life changes, or understand how others relate to them. The practitioner may be the only person to view the test results; for example, in the case of making a decision as to a client's status or state. That decision may be used to help make a decision to admit a person for treatment or to refer that person to another agency or practice.

Tests Used in Industry

One of the more fascinating areas of testing is that of selection, progression, and promotion in industry. In this setting, there are many different stakeholders, as well as federal, state, and sometimes local regulations and requirements that compete with psychometric characteristics of the test.

In the workplace setting, the purpose of testing is to determine the best candidate for a specific position or job. The goal is to determine the specific knowledge, skills, and abilities needed to be successful in that position and to measure as many of these as is possible prior to hiring, training, or promoting an individual. In industry, hiring a worker is associated with enormous costs, including wages, relocation, training, and benefits. Making a poor choice may have devastating effects on an organization and can develop into a health or safety issue, depending on the industry and specific job.

Many of the tests used in industry are specific to the company, plant site, and job. Developed by outside consultants or in-house personnel, these tests utilize job and task analysis to develop the content of the test and determine the appropriate level of knowledge, skill, and ability needed. This process can be very costly. Hence, firms must engage in a cost analysis to determine whether building or buying a test will benefit the company. Typically, this cost analysis looks for

savings in training time, error rates, employee turnover, and other factors in determining the benefit to the company.

Conclusion

Any given test may be classified and used in many ways. The practitioner has a responsibility to look at the testing situation, the decisions to be made by each of the stakeholders in that situation, and the available test instruments in order to determine the best course of action. *Measuring Up* provides insights into many of the issues encountered in the testing arena and provides practitioners with guidance and resources to help them do their work. Many other books are available that review or critique commercially available tests. In addition, several professional organizations address issues of testing, measurement, and assessment. The newsletters and journals of these organizations can provide information beneficial in understanding how a test can be used. You can find specific resources and references to these in chapter 53.

It is important to understand the nature of tests and how they may be used and classified. It is more important, however, to use the best tools available, acquire the training necessary to use these tools correctly, then make good conservative use of the test results in light of the setting and the individuals involved.

References

AERA, APA, & NCME. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory.* Belmont CA: Wadsworth.

Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: Macmillan.

DuBois, P. H. (1970). *A history of psychological testing.* Boston, MA: Allyn & Bacon.

Hood, A. B., & Johnson R. W. (1997). *Assessment in counseling: A guide to the use of psychological assessment procedures.* Alexandria, VA: American Counseling Association.

Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation* (8th ed.). Needham Heights, MA: Allyn & Bacon.

Janda, L. H. (1998). *Psychological testing: Theory and applications.* Needham Heights, MA: Allyn & Bacon.

Murphy, L. L., Conoley, J. C., & Impara, J. C. 1994 . *Tests in print IV.* Lincoln: University of Nebraska Press.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[ ] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[X] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").